Т. 85, № 6 ЖУРНАЛ ПРИКЛАДНОЙ СПЕКТРОСКОПИИ НОЯБРЬ — ДЕКАБРЬ 2018
V. 85, N 6 JOURNAL OF APPLIED SPECTROSCOPY NOVEMBER — DECEMBER 2018

DYNAMIC SPECTRUM EXTRACTION METHOD BASED ON ABSOLUTE DIFFERENCE SUMMATION AND STATISTICAL THEORY

G. Li, H. L. Wang*, M. Zhou, Y. Peng, L. Lin

Tianjin University, 92 Weijin Road, Nankai District, Tianjin 300372, China; email: lelestudent@tju.edu.cn

Dynamic spectrum theory has great importance in the field of noninvasive measurement of blood components. To enhance the computational efficiency and data utilization of the existing extraction methods, this paper proposes a new one based on the absolute difference summation (ADS) and statistical theory. The ADS is used to obtain the eigenvalue from a photoplethysmography (PPG) signal. The statistical method is used to obtain the final dynamic spectrum. The experimental data of PPG signal from 133 volunteers were extracted by the new method and the single-trial (ST) extraction method, and the partial least squares model was used to build the calibration models. Compared with the ST extraction, the new method showed better prediction ability. The correlation coefficient of the prediction set increased from 0.85 to 0.92, and the root mean square error of the prediction decreased from 13.49 to 9.86 g/L, which proved that this method can significantly improve the quality of the dynamic spectrum.

Keywords: *noninvasive measurement of blood components, photoplethysmography, dynamic spectrum, absolute difference summation, modeling accuracy.*

МЕТОД ПОЛУЧЕНИЯ ДИНАМИЧЕСКОГО СПЕКТРА, ОСНОВАННЫЙ НА СУММИРОВАНИИ АБСОЛЮТНОЙ РАЗНОСТИ И СТАТИСТИЧЕСКОЙ ТЕОРИИ

G. Li, H. L. Wang*, M. Zhou, Y. Peng, L. Lin

УДК 535.33:611-018.5

Тяньцзиньский университет,
шоссе Вэй Цзинь, 92, Нанькай, Тяньцзинь, 300372, Китай; email: lelestudent@tju.edu.cn

(Поступила 4 декабря 2017)

Предлагается метод получения динамического спектра, основанный на суммировании абсолютной разности (САР) и статистической теории. САР используется для определения собственного значения по сигналу, полученному с помощью фотоплетизмографии (ФПГ). Для получения конечного динамического спектра применяется статистический метод. Экспериментальные данные ФПГ-сигнала для 133 добровольцев обработаны с помощью нового метода и метода однократной экстракции, а для построения калибровочных моделей использован метод частных наименьших квадратов. По сравнению с однократной экстракцией новый метод показал лучшую способность предсказания. Коэффициент корреляции прогнозного набора увеличился с 0.85 до 0.92, а среднеквадратическая погрешность предсказания уменьшилась с 13.49 до 9.86 г/л. Это доказывает возможность метода значительно улучшить качество динамического спектра.

Ключевые слова: *неинвазивное измерение компонентов крови, фотоплетизмография, динамический спектр, суммирование абсолютной разности, достоверность моделирования.*

Introduction. In recent years, noninvasive detection of blood components has become a hot spot in the biomedical engineering field [1–4]. Among the related methods, the near-infrared spectroscopy technology has attracted much attention because of its advantages, such as rapid processing, good specificity, and large amount of information, and has achieved significant research progress [5, 6]. Jeon proposed to establish a calibration model by measured spectra data to analyze hemoglobin concentration [7]. Kraitl developed the sensor design and the system construction, but the spectral algorithm had a low performance largely due to the lack of source information on the spectral data [8]. Yamakoshi extracted the ratio of the alternating component to the direct one in the PPG signal waveform to obtain effective information from the pulsating section of the arterial blood and used the experimental equipment with high performance [9], but this method used less available data compared to noise and external disturbance.

As can be seen from the above studies, the extraction of valid data is crucial to spectral analysis. It is a key issue to overcome the individual differences and the influence of the non-pulsating part of human tissue. Li Gang put forward the theory of DS in 2004 aiming at this problem [10]. This theory led to a series of scientific research results in many fields, such as extraction of spectral data, pretreatment of spectral data, and the model establishment and validation [11–14]. In fact, the essence of this theory is similar to the time-resolved spectroscopy proposed by Yamakoshi [15] and the differential spectral theory proposed by Chen Xingdan [16]. However, in the latter two theories, the single extraction method reduces the measurement accuracy. So the spectrum extraction is an extremely important link in the signal transmission. According to the DS theory, the spectrum eigenvalue is the difference of the maximum and minimum absorbance values in a single PPG waveform period [17].

Frequency domain analysis and ST estimation are currently the two main types of extraction methods for DS. In the frequency domain, the DS value is formed by the proportional relation of the fundamental harmonic amplitude of the PPG signal with its Fourier transform [18–20]. However, this method can easily be affected by abnormal pulse noise. In the time domain, the ST extraction method [21] was proposed. It uses a statistical method to get rid of gross errors [22, 23] and improve the quality of the initial DS.

In this paper, a DS extraction method based on the ADS and statistical theory is proposed. It makes up for the defects of the method above-mentioned and significantly improves the accuracy of the extraction.

DS theory. Using near infrared and visible light, we can obtain a PPG signal from the tip of fingers. The DS value is composed of the change in the absorbance of blood volume by the sequential extraction of the logarithm PPG signal at each wavelength [10, 24]. It is theoretically proved that the variation of the transmitted light intensity is only related to the filling and contraction of arterial blood. Other body tissues remain almost unchanged in the course of pulse fluctuations. Therefore, DS can eliminate the effect of this part of the tissue [17].

As mentioned above, the basic idea is to use the peak value of the logarithm PPG at each wavelength to extract the absorption spectrum of pulsating arterial blood, which is called the peak-value extraction method. In the actual test, the limited sampling rate will lead to a large error in the waveform peak sign of the logarithm PPG signal, and it can be easily affected by noise or external interference, resulting in poor accuracy of the model.

DS time-domain ST extraction method. The ST extraction method uses the superposition average effect of logarithmic photoelectric pulse wave at all wavelengths to obtain the log PPG template. It uses the effective rising edge of this template to eliminate errors in the corresponding PPG waveform, calculates the absorbance difference value, and then gets a DS value [21]. In this way, it not only eliminates the DS values with gross errors, but also lets the spectral data be collected at the same time, which reduces the influence of abnormal interference waveform on the quality of the DS line. As we can see, for the signal polluted by jitter and pressure change, the ST method has stronger ability to restrain the noise. However, the disadvantage of this method is that the implementation process is more cumbersome for a large number of data samples, and this method still has not achieved the desired results.

DS extraction method based on ADS and statistical theory. According to DS, the amplitude difference between any two points of the PPG signal can reflect the absorbance of blood. By superimposing these differences, the differential spectrum can be obtained. However, the simple algebraic summation may lead to the plus-minus elimination and lose their efficacy. So, this paper uses the absolute value of these differences to avoid this problem. Using these absolute differences as the sub DS and calculating the stack average of all of them may lead to some singular value in the spectral data and reduce its SNR. So, this paper chooses using the piecewise summation to obtain the sub DS. In order to avoid the low utilization rate of raw data, the new method adopts the data overlapping partition, and the statistical method is used to optimize an effective sub DS to further suppress the impact of gross error on the DS data accuracy.

In conclusion, the new method not only avoids the impact of random pulse error, but also makes full use of the collected PPG data by statistical methods. This method also has the advantages of the above methods, such as the high computational efficiency. According to the experiment below, this method can initially improve the dynamic spectrum SNR.

Experimental device and data acquisition. The device used to collect PPG signals is mainly composed of a light source, a spectrometer, and a PC machine, as shown in Fig. 1. This experiment uses a Philips bromine tungsten lamp as the light source, with wavelength ranging from 463 to 1356 nm. The spectrometer type is AvaSpec-HS1024x58TEC from Avantes. The spectrometer covers an effective wavelength range from 200 to 1200 nm with a resolution of better than 1.2 nm and SNR of 1000:1. The light source illuminates the subjects' fingers directly. Then the light transmitted from the finger is sent to the spectrometer through the fiber. Finally, the spectral data collected by the spectrometer is delivered to the PC machine for data storage and subsequent processing.

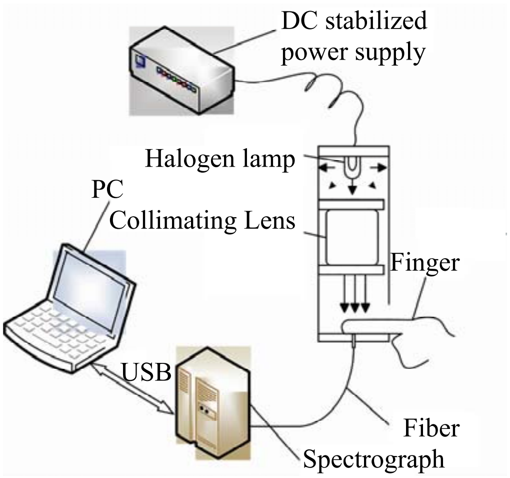

Fig. 1. The acquisition equipment of PPG.

The participants gently cover the fiber's optic probe and keep the contact pressure basically constant. This experiment sets integral time of the spectrometer as 20 ms and the synchronous acquisition of the PPG signal is completed at 28 s. So, 1400 sampling points for each wavelength were obtained.

To improve the original data SNR, only the wavelength band of the larger transmitted intensity can participate in the calculation. So in this experiment, only the PPG data in the ~613–1112 nm band are chosen for processing. Figure 2 shows the PPG signal of one volunteer.

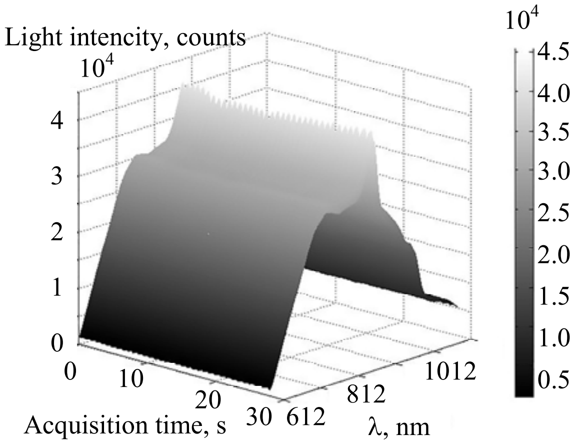

Fig. 2. The PPG signal of one volunteer.

The data processing steps of the new method. Based on the principle of the DS theory and the new extraction method, we designed the following steps to extract the spectrum:

(1) Carry out simultaneous acquisition of full-band photoelectric volume pulse waves and get them logarithmically transformed. We assumed the number of wavelengths, N, to be 500, the length of window, W, to be 100, and the sliding step length, S, to be 4.

(2) Use the Dual-tree Complex Wavelet Transform (DTCWT) [25] to process the logarithmic PPG signal within the wavelength band mentioned above, which is in fact the processing of a smoothing filter for the original signal. Its validity for the source signal pretreatment of the DS data has been proved in the relevant literature [26].
(3) We assumed the number of sample points, M, to be 1400. Taking one of the wavelengths as an example, the first window of the logarithmic PPG signal can be obtained from its first W sampling points. This method calculates the logarithm amplitude difference between each two adjacent sampling points in the window and then gets its absolute value. Finally, W-1 absolute differences can be obtained. Then the window is slid for S sampling points sequentially. As mentioned above, other W–1 absolute differences are obtained. The rest can be carried out in the same manner. In summary, several windows of the logarithmic PPG signal can be finally obtained. The windows' number is recorded as N_w.
(4) Accumulate the W–1 absolute differences which are inside of each window. An ADS sequence will be obtained, whose length is N_w, the same as the number of windows.
(5) According to Fig. 2, it can be seen that the waveform at different wavelengths has good consistency. Therefore, this experiment gets the superposition average of ADS in the same position of sequences at different wavelengths to obtain a template. It gets the values of the ADS sequence corresponding divided by the value in the template and obtains a set of ratio coefficients recorded to K_i (i = 1, 2, 3, …, N_w). Further, it superimposes all the coefficients K_i and gets the average value, the normalized coefficient $\bar{K}$. With the proportional coefficient K_i at each wavelength divided by $\bar{K}$, the aplanatism normalization of DS can be made.
(6) The distance between each normalized ADS is recorded as X^i, and the template of the normalized ADS is recorded as $\bar{X}$. The Euclidean distance, $D(X^i, \bar{X})$, is used to describe their similarity. In this operation, $D(X^i, \bar{X}) = \sqrt{\sum_{\lambda=1}^{N} | X_\lambda^i - \bar{X}_\lambda |}$. This experiment then calculates the average Euclidean distance $\bar{D}$, residual error ν_i, and the standard deviation σ as

$$\bar{D} = \frac{1}{N}\sum_{i=1}^{N} D\left(X^i - \bar{X}\right),\ \nu_i = D(X^i - \bar{X}) - \bar{D},\ \sigma = \sqrt{\sum_{i=1}^{N} \nu_i^2 \Big/ (N-1)}\ .$$

If $|\nu_i| > 2\sigma$, it is considered that this ADS has a gross error and should be removed. Otherwise it will be retained. After all the gross errors are eliminated, the normalized ADS sequence is finally obtained.
(7) Get the superposition average of the normalized ADS sequences as the DS values. This experiment arranges the DS values according to the wavelength from small to large, as shown in Fig. 3.

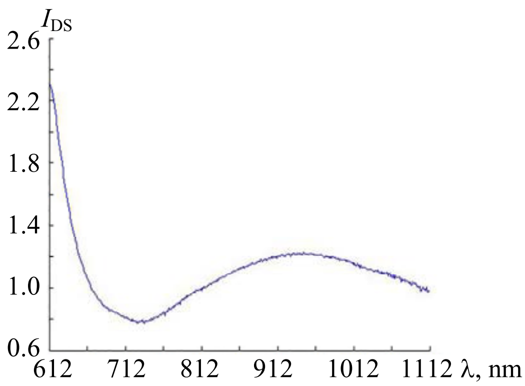

Fig. 3. The waveform of DS from the sample No. 107.

Results and discussion. *Modeling method of PLS.* In this experiment, the 133 sets of data collected by the experimental device are adopted for the DS extraction. According to the different methods of data processing, the new extraction method and the ST extraction method are used as the experimental strategy and the control strategy, respectively. In order to evaluate the two methods, this paper uses the PLS method to model the spectral data obtained from the two methods and uses the hemoglobin concentration of the samples as reference value.

Since the accuracy of the model is directly related to the correlation and the error magnitudes between the DS and reference values, the effectiveness and stability of the DS extraction process can also be indirectly reflected. The reference values of this experiment are obtained by blood gas analyzers from the volunteers' blood samples collected from arterial blood in the same state as at the spectral data measurement. Therefore there is no error between the organization systems.

Comparison of modeling results for two methods. In this experiment, the general information of these volunteers is listed in Table 1. In order to improve the original signal's SNR in modeling, the data of each sample's DS is processed by a Savitzky–Golay filter (SG filter) with 5 points and 1 order. The model set and prediction set are selected by the non-uniform extraction. Finally, 19 samples were selected as the prediction set, and the remaining 114 samples were used as the model set. The successive elimination and cross validation method is used to choose the principal component number. This paper sets the correlation coefficient of the prediction set and RMSEP as the main indicators of the model evaluation.

The main indicators of the modeling evaluation are listed in Table 2. It can be seen that when the number of the principle components is 14, the correlation coefficient of the ST extraction method reaches the maximum value, 0.8496, and RMSEP reaches the minimum value, 13.4869 g/L. When the number of the principle components is 15, the correlation coefficient of the experimental strategy reaches the maximum value, 0.9186, and RMSEP reaches the minimum value, 9.8613 g/L.

TABLE 1. The General Information of the Subjects

Total sample size	Gender	Sample size	Age		Hemoglobin concentration, g/L	
133	Male	70	Max	87	Max	169
			Min	22	Min	110
			Average	45.69	Average	145.8
	Female	63	Max	80	Max	158
			Min	24	Min	86
			Average	43.28	Average	126

TABLE 2. The Evaluation of the Models

Method	Number of principle component	Correlation coefficient		Root mean square error, g/L	
ST method	14	prediction	0.8496	prediction	13.4869
		calibration	0.9215	calibration	6.8821
New method	15	prediction	0.9186	prediction	9.8613
		calibration	0.9607	calibration	5.9512

Figure 4 shows the effect of the two methods on the prediction of hemoglobin concentration of the sample set. It can be seen that the study group was significantly better than the control group. Compared to the ST extraction method, the new extraction method has a higher prediction correlation coefficient and a lower RMSEP. The method heightens the modeling accuracy and significantly improves the efficiency and stability of DS.

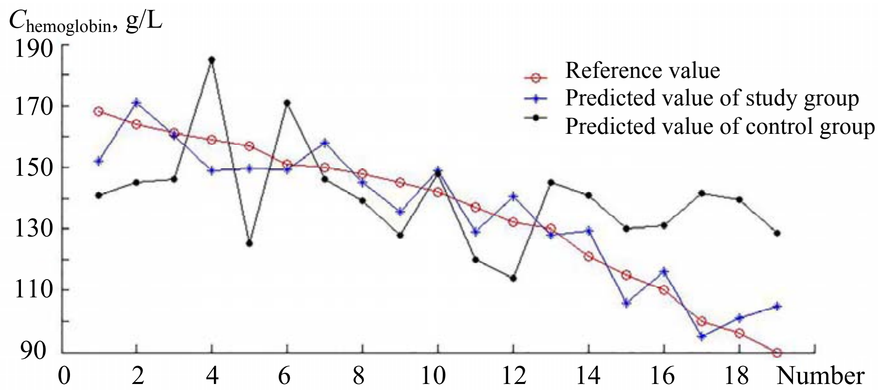

Fig. 4. The prediction results of the two methods.

Conclusion. Aiming at the shortcomings of the time-domain ST extraction method, such as process complexity and low effective data utilization, a new extraction method using ADS and statistical theory is proposed. The DS data from 133 volunteers were obtained respectively using the new method and the ST extraction method. It can be seen that, compared to the ST extraction method, the new extraction method has a higher prediction correlation coefficient and a lower RMSEP. This method increases the modeling accuracy and significantly improves the efficiency and stability of the DS. Since the method is simple and fast, it is more suitable for the noninvasive detection of blood components in the microprocessor application. This method will be helpful to promote the application of the DS theory.

Acknowledgment. This study was supported by the Tianjin Application Basis & Front Technology Study Programs (No. 14JCZDJC33100).

REFERENCES

1. J. W. Mcmurdy, G. D. Jay, S. Selim, G. Crawford, *Clin. Chem*., **54**, 264–272 (2008).
2. Y. Uwadaira, N. Adachib, A. Ikehatac, S. Kawanoa, *J. Near Infrared Spectrosc.*, **18**, 291–300 (2010).
3. G. Bertrand, M. Drame, C. Martageix, C. Kaplan, *Blood*, **117**, 3209–3213 (2011).
4. H. Wang, H. Q. Xu, L. M. Qu, X. M. Wang, R. H. Wu, X. Z. Gao, Q. L. Jin, J. Q. Niu, *Eur. J. Gastroenterol. Hepatol.*, **28**, 997–1002 (2016).
5. K. Maruo, M. Tsurugi, J. Chin, T. Ota, *Select. Top. Quantum Electron.*, **9**, 322–330 (2003).
6. A. Enejder, S. G. Sasic, M. Feld, *J. Biomed. Opt.*, **10**, 885–889 (2005).
7. K. J. Jeon, S. Kim, K. K. Park, J. Kim, G. Yoon, *J. Biomed. Opt.*, **7**, 45–50 (2002).
8. J. Kraitl, U. Timm, E. Lewis, H. Ewald, *Proc. SPIE*, **7572**, 757209 (2010).
9. Y. Yamakoshi, M. Ogawa, T. Yamakoshi, T. Tamura, K. Yamakoshi, *31st Annual Int. Conf. IEEE EMBS*, Minneapolis, Minnesota, USA, 126–129 (2009).
10. G. Li, Y. L. Liu, L. Lin, Y. Wang, *3rd Int. Symposium on Instrumentation Science and Technology*, August 18–22, Xi'an, China, 3-0875-0880 (2004).
11. G. Li, Y. Wang, Q. X. Li, X. X. Li, L. Lin, Y. L. Liu, *J. Infrared Millimeter Waves*, **25**, 345–348 (2006).
12. G. Li, Y. L. Liu, L. Lin, Y. Wang, *Opt. Precision Eng*., **14**, 816–821 (2006).
13. G. Li, H. Q. Wang, Z. Zhao, L. Lin, *Spectrosc. Spectr. Anal.*, **30**, 2385–2389 (2010).
14. L. Lin, Q. R. Zhan, M. Zhou, S. J. Xu, G. Li, *RSC Advances*, **6**, 113322–113326 (2016).
15. Y. Yamakoshi, M.Ogawa, T. Yamakoshi, M. Satoh, M. Nogawa, S. Tanaka, T. Tamura, P. Rolfe, K. Yamakoshi, *Int. Conf. IEEE Engineering in Medicine & Biology Society*, 2964–2967 (2007).
16. Y. H. Luo, Z. Chen, X. D. Chen, *Opt. Precision Eng.*, **16**, 759–763 (2008).
17. H. Q. Wang, G. Li, Z. Zhao, L. Lin, *Trans. Inst. Measure. Control*, **35**, 16–24 (2013).
18. G. Li, X. X. Li, L. Lin, Y. L. Liu, Y. Wang, *Spectrosc. Spectr. Anal.*, **26**, 263–266 (2006).
19. G. Li, Q. X. Li, L. Lin, X. X. Li, Y. Wang, Y. L. Liu, *Spectrosc. Spectr. Anal.*, **26**, 2177–2180 (2006).
20. Y. Wang, G. Li, L. Lin, Y. L. Liu, X. X. Li, *Spectrosc. Spectr. Anal.*, **27**, 91–94 (2007).
21. G. Li, C. Xiong, H. Q. Wang, L. Lin, B. J. Zhang, Y. Tong, *Spectrosc. Spectr. Anal.*, **31**, 1857–1861 (2011).
22. G. Li, M. Zhou, L. Lin, *Opt. Quantum Electron.*, **46**, 691–698 (2014).
23. W. Q. He, X. X. Li, M. J. Wang, G. Li, L. Lin, *Anal. Methods*, **7**, 5565–5573 (2015).
24. G. Li, Y. Wang, L. Lin, Y. L. Liu, X. X. Li, S. Y. Lu, *IEEE Conference Publications-2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 1960–1963 (2005).
25. I. W. Selesnick, R. G. Baraniuk, N. C. Kingsbury, *IEEE Signal Process. Mag.*, **22**, 123–151 (2005).
26. Y. Peng, G. Li, M. Zhou, H. L. Wang, L. Lin, *RSC Adv.*, **7**, 11198–11205 (2017).